

X IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA

v.

ROBERT BOWERS

)
)
)
)
)

Criminal No. 18-292

**MOTION TO EXCLUDE FIREARM TOOLMARK EVIDENCE, AND REQUEST
FOR *DAUBERT* HEARING AND FOR DISCOVERY IN SUPPORT OF MOTION**

1001 Liberty Avenue
Suite 1500
Pittsburgh, PA 15222
(412) 644-6565

Judy Clarke
Clarke Johnston Thorp & Rice, PC

Michael J. Novara
First Assistant Federal Public Defender

Elisa A. Long
Supervisory Assistant Federal Public
Defender

Attorneys for Defendant,
Robert Bowers

TABLE OF CONTENTS

I. INTRODUCTION.....	5
II. THE FIREARM TOOLMARK EVIDENCE IN THIS CASE.....	7
III. ARGUMENT	8
A. DAUBERT REQUIRES RIGOROUS EVALUATION AND REEVALUATION OF THE RELIABILITY OF PARTICULAR METHODOLOGIES AND HOW THEY ARE APPLIED.	11
B. THREE COMMITTEES OF EXPERTS FROM THE SCIENTIFIC COMMUNITY CONDUCTED A COMPREHENSIVE REVIEW OF THE EMPIRICAL DATA AND CONCLUDED THAT FIREARMS IDENTIFICATION IS NOT SCIENTIFICALLY VALID..	15
1. The 2008 NRC Ballistics Imaging Report	17
2. The 2009 NRC Forensics Report	17
3. The 2016 PCAST Report	20
C. THIS COURT SHOULD EXCLUDE FIREARMS IDENTIFICATION EVIDENCE AS SCIENTIFICALLY INVALID.....	30
D. THE COURT SHOULD EXCLUDE ANALYST MILLS’ TESTIMONY BECAUSE THERE IS NO WAY TO DISCERN WHETHER HIS RESULTS ARE RELIABLE.....	34
1. Proficiency Testing is Insufficiently Rigorous to Ensure Reliability.	36
2. Without further discovery of the litigation packets, it is impossible to determine (i) whether any protocols that the examiner applied and the documentation produced are sufficient to ensure reliable and repeatable analyses and (ii) whether those protocols were in fact followed.	38
E. THE PREJUDICIAL IMPACT OF FIREARMS ANALYSIS EVIDENCE SUBSTANTIALLY OUTWEIGHS ANY PROBATIVE VALUE.	41
F. INTRODUCTION OF INDIVIDUALIZATION TESTIMONY WOULD VIOLATE MR. BOWERS’ DUE PROCESS RIGHTS.	43
G. IN THE ALTNERATIVE, THIS COURT SHOULD LIMIT THE SCOPE OF THE EXPERT’S TESTIMONY, REQUIRE ERROR RATE TESTIMONY, AND EXCLUDE ANY CONCLUSIONS UNGROUNDED IN A RELIABLE SCIENTIFIC FOUNDANTION..	44
IV. CONCLUSION.....	47

IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA)	
)	
v.)	Criminal No. 18-292
)	
ROBERT BOWERS)	

**MOTION TO EXCLUDE FIREARM TOOLMARK EVIDENCE,
AND REQUEST FOR A *DAUBERT* HEARING AND FOR DISCOVERY IN
SUPPORT OF MOTION**

Defendant Robert Bowers, through counsel, moves the Court for an order, following an evidentiary hearing, excluding the government's firearm toolmark evidence in his case. Mr. Bowers makes this motion pursuant to the Federal Rules of Evidence Rules 104(a), 402, 403, 702, and 703; Federal Rule of Criminal Procedure Rule 16; 18 U.S.C. § 3593(c); and the Fifth, Sixth, and Eighth Amendments to the United States Constitution.

The grounds for this motion are that:

- there is no reliable scientific basis for the identification claims and conclusions regarding the firearm toolmark examinations in this case, and thus this testimony is inadmissible under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), and *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999);
- the testimony is inadmissible under the 2000 amendments to Rule 702 and Rule 703 in that (a) the testimony is not based upon sufficient facts or data,

- (b) the testimony is not the product of reliable principles and methods, and
- (c) the individuals who performed the firearm toolmark examinations have not applied the principles and methods reliably to the facts of the case;
- the specific testimony based on firearm toolmark analysis offered in this case is irrelevant and inadmissible under Rule 402; and
- any weak probative value of the proposed testimony is also substantially outweighed by the danger of unfair prejudice, confusion of the issues, and misleading the jury, and by considerations of undue delay, waste of time, and needless presentation of cumulative evidence and is thus inadmissible under Rule 403 and the due process, fair trial, and cruel and unusual provisions of the Constitution and 18 U.S.C. § 3593(c).

In addition, and preliminary to ruling on the admissibility of the firearm toolmark evidence, Mr. Bowers asks this Court to compel the government to provide forthwith the “case litigation” packets that support the findings and conclusions of FBI Firearm/Toolmarks Unit Casework Analyst Brett A. Mills 41-page Firearm Toolmark Report (“Firearm Toolmark Report”).¹ The case litigation packets, as well as additional information that the defense may need to request after an initial review of the litigation packets, are essential for the defense to determine, through independent expert analysis, whether there are grounds in addition to those raised in this motion for exclusion of the government’s firearm toolmark claims and conclusions in this case. Identifying the

¹ The Firearm Toolmark Report is attached as Exhibit 1.

proper experts and narrowing the *Daubert* issues require the production of the case litigation packet materials.

At the status hearing held on November 2, 2020, the defense noted the complexity of *Daubert* issues surrounding the firearm toolmark (as well as DNA) evidence and explained that this was an area where the defense could make progress in this case despite the COVID delays. Transcript of November 2, 2020 Telephone Conference at 4-5. The defense explained that the government's provision of the case litigation packets would facilitate the *Daubert* litigation, and the Court indicated that it would entertain a written request to that effect. *Id.* at 7-8, 19 ("I'm going to wait and see what the defense files in terms of the *Daubert* motions.").

In accordance with the Court's instructions, Mr. Bowers brings the present motion, which outlines a number of *Daubert* issues that can be raised in this case based on Analyst Mills' conclusory report alone. Necessarily, however, this preliminary *Daubert* motion is not exhaustive and must be supplemented in the future since any flaws in the underlying analysis of the firearm toolmark claims and conclusions as applied to this specific case are impossible to fully understand without access to the actual testing observations, analysis, and other pertinent information that the defense may need to request after review of those materials.

I. INTRODUCTION

There is a sea change underway in the scientific and judicial evaluation of firearm toolmark identification claims and conclusions. Until recently, federal and state courts around the United States have permitted, without much scrutiny, the testimony of

firearms toolmark identification experts as “scientific” proof that the bullets or shell casings recovered from a crime scene were fired from a particular gun. These claims of identification have been based on a belief that due to “matching” microscopic toolmarks left on spent ammunition (as a result of having been fired or expelled), an expert can determine whether the same firearm fired or expelled two or more pieces of ammunition to the exclusion of any other firearm. Indeed, prosecutors have relied on this testimony to secure convictions nationwide for over a century.

It has become increasingly clear, however, that the methodology of firearm toolmark pattern matching is not based on reliable principles and methods and thus lacks foundational validity. Three comprehensive reports authored by independent committees of scientists in 2008, 2009, and 2016—two issued by different subcommittees of the National Research Council (NRC)² and one issued by the President’s Council of Advisors on Science and Technology (PCAST)—have rejected the claim that firearms identification is valid and reliable science. The most recent 2016 report, issued following a review of more than 2,000 articles and presentations by members of the forensic community, states that firearms identification “falls short of the scientific criteria for foundational validity.” President’s Council of Advisors on Science and Technology,

² The National Research Council (“NRC”) is a component of the National Academy of Sciences, which was created by congressional charter in 1863 to “investigate, examine, experiment, and report upon any subject of science.” ACT TO INCORPORATE THE NATIONAL ACADEMY OF SCIENCES, § 3, 12 Stat. 806 (1863), <http://www.nasonline.org/about-nas/leadership/governing-documents/act-of-incorporation.html> (last visited Jan. 8, 2021). The NRC was established in 1916 “to associate the broad community of science and technology with the Academy’s purposes of furthering knowledge and advising the federal government.” National Research Council, BALLISTIC IMAGING iii (2008) (“NRC 2008 Ballistic Imaging Report”). There are not more independent or prestigious scientific organizations in the United States.

FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING VALIDITY OF FEATURE-COMPARISON METHODS 11 (2016) (“PCAST Report”).³

Courts, accordingly, have taken notice of these important scientific reports and increasingly are excluding or limiting firearm toolmark identification claims due to their lack of reliability or validity.

II. THE FIREARM TOOLMARK EVIDENCE IN THIS CASE

The conclusory report of FBI Analyst Brett Mills provided through discovery indicates that [REDACTED]

[REDACTED]. The report [REDACTED]
[REDACTED].

The unreliable and invalid “identification” claims purport to [REDACTED]
[REDACTED]
[REDACTED], presumably through the observation of
microscopic markings. In addition, the report indicates [REDACTED]
[REDACTED]
[REDACTED]

³ The PCAST report is available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (last visited Jan. 8, 2021).

The government alleges that Mr. Bowers used multiple firearms to kill and injure multiple victims in this case. Firearm-related items were subjected to visual examination and compared against one another for the presence of general class characteristics and microscopic markings.

Based on the results of the Firearm Toolmarks analysis performed by Analyst Mills, he came to certain conclusions regarding [REDACTED]. These “identification” claims to a particular firearm are inherently unreliable, invalid, and more prejudicial than probative as evidence in this case.

III. ARGUMENT

The FBI Laboratory, like other laboratories with firearm toolmark units across the United States, continues to make claims of “identification” or “association” despite growing criticism over the scientific foundational validity of such claims. Distinguished scientific bodies such as the National Academy of Sciences and the President’s Council of Advisors on Science and Technology have specifically found that these claims are not supported by rigorous scientific studies and are yet to be established as reliable.

As a threshold matter, an application of Rule 702’s scientific validity and reliability requirements warrants the exclusion of Analyst Mills’ opinions. Unless and until the government can show that firearms examiners have produced appropriate, independent empirical evidence: (1) that firearms have unique or individual characteristics; (2) that firearms consistently imprint those characteristics on bullets and shell casings of the type at issue in this case; and (3) that examiners can identify and

evaluate those marks using a reliable and repeatable method, with a known and acceptable error rate, the proposed evidence cannot be admitted at trial. The government cannot meet this burden, as multiple independent scientific bodies have concluded over the past decade.

The National Academy of Sciences and PCAST scientific reports made several findings compelling the conclusion that this Court should not admit Analyst Mills' proffered testimony in this case:

- The theory of firearms identification is “not a scientific theory,” but rather a “claim that examiners applying a subjective approach can accurately individualize the origin of a toolmark.” PCAST Report at 60.
- “The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been demonstrated.” National Research Council, *BALLISTIC IMAGING 3* (The National Academies Press), available at <https://doi.org/10.17226/12162> (last visited Jan. 8, 2021) (“Ballistics Imaging Report”).
- “A fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process.” National Research Council, Committee on Identifying the Needs of the Forensic Science Community, *STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD* 154 (2009) (“NRC Forensics Report”).
- The method is subjective, with examiners using their personal judgment to select which features to compare; and the reasoning employed to reach a conclusion is circular—a match can be declared when there is “sufficient agreement,” and when there is sufficient agreement is based on the personal judgment of each examiner. PCAST Report at 47, 60, 104, 113.
- Because the method is “subjective,” foundational validity and reliability “can *only* be established through multiple independent black-box studies” and “at present there is only a single study that was appropriately designed” PCAST Report at 106, 111.
- As a result, firearms analysis currently “falls short of the scientific criteria for foundational validity.” PCAST Report at 111.

These three scientific reports authored by separate committees of nationally recognized scientists found that the “fundamental assumptions” underlying firearms examination have not been demonstrated; that the theory is “not a scientific theory”; that the method is “subjective” and not well defined; and that there is “insufficient empirical evidence” establishing validity and estimating reliability. *See United States v. Davis*, No.18-cr-00011, 2019 WL 4306971, at *5 (W.D. Va. Sept. 11, 2019) (“The series of reports issued on this subject [firearms examination] reflects the ever-growing number of members of the legal and scientific communities who recognize problematic aspects of this discipline.”).

Exclusion is the only appropriate remedy in this case, as cross-examination cannot cure the prejudice from the admission of evidence derived from a method that has not been proven valid and that lacks estimates of its reliability. Scientific expert testimony carries with it the “aura of special reliability and trustworthiness,” creating a grave risk that jurors will receive it without a critical eye. *Motorola Inc. v. Murray*, 147 A.3d 751, 753 (D.C. 2016) (en banc) (quoting *Ibn-Tamas v. United States*, 407 A.2d 626, 632 (D.C. 1979)). Perhaps because juries may view forensic testimony with a less-than-critical trust, the use of unreliable forensic science is one of the leading causes of wrongful convictions. *See* Brandon L. Garrett, *Judging Innocence*, 108 Colum. L. Rev. 55, 81–88 (2008).

Absent granting the relief of exclusion based on the pleadings, the Court must hold an evidentiary hearing on the issue. At a bare minimum, the Court should limit the

testimony of the government's firearms examiner to what was done and observed and exclude any testimony that purports to claim any degree of scientific validity to the methodology used or that an "identification" to a particular firearm can be made. The examiner's conclusions should be limited to *including* or *excluding* a weapon at issue as a potential source of the crime scene evidence; any individualization, association, or probabilistic claims are not supported by the science.

A. *DAUBERT* REQUIRES RIGOROUS EVALUATION AND REEVALUATION OF THE RELIABILITY OF PARTICULAR METHODOLOGIES AND HOW THEY ARE APPLIED.

As this Court is aware, the standard for the admission of expert testimony is Rule 702 of the Federal Rules of Evidence, as interpreted by *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993), and its progeny.⁴ Federal Rule of Evidence 702 provides as follows:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;
- (c) the testimony is the product of reliable principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case.

Fed. R. Evid. 702.

⁴ See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999); *General Electric Co. v. Joiner*, 522 U.S. 136 (1997). The *Kumho Tire*, *Joiner*, and *Daubert* cases have come to be known as the "Daubert Trilogy."

This “‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition of admissibility.”” *See Davis*, 2019 WL 4306971, at * 3 (quoting *Daubert*, 509 U.S. at 591–92). In applying Rule 702, a trial court must act as a gatekeeper and make “a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” *Daubert*, 509 U.S. at 592–93.⁵

The Court in *Daubert* suggested factors a trial court may consider in applying Rule 702: whether the theory or technique has been tested; whether it “has been subjected to peer review and publication;” “the known or potential rate of error;”⁶ and “the existence and maintenance of standards controlling the technique’s operation.” *Daubert*, 509 U.S. at 593–94. No one of these factors is individually controlling, and additional factors may apply in a particular case. *Id.* Furthermore, the *Daubert*/Rule 702 test applies not just to novel scientific evidence or to unconventional techniques, but to all scientific or technical methods. As the *Daubert* Court stated, a “trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.” *Daubert*, 509 U.S. at 589.

⁵ *Daubert*’s “general holding—setting forth the trial judge’s general ‘gatekeeping’ obligation—applies not only to testimony based on ‘scientific’ knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge.” *Murray*, 147 A.3d at 755 (quoting *Kumho Tire Co.*, 526 U.S. at 141).

⁶ The collective experience of examiners in the field and in courts, no matter how extensive, is not a scientific basis to demonstrate validity and reliability. Claims to the contrary have been soundly rejected by the scientific community. “Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy.” PCAST Report at 46.

Daubert and Federal Rule of Evidence 702 have been consistently applied in the Third Circuit, requiring trial court judges to act as “gatekeepers” to ensure that “any and all expert testimony or evidence is not only relevant, but also reliable.” *Pineda v. Ford Motor Co.*, 520 F.3d 237, 243 (3d Cir. 2008) (quoting *Kannankeril v. Terminix Int’l, Inc.*, 128 F.3d 802, 806 (3d Cir. 1997) (citing *Daubert*, 509 U.S. at 589)); see *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717 (3d Cir. 1994). In *Schneider ex rel. Estate of Schneider v. Fried*, the Third Circuit provided what is referred to as a shorthand for the three-part test that must be satisfied in order for an expert to testify: a test requiring the Court to look at the qualification, reliability, and fitness of the expert. 320 F.3d 396, 404 (3d Cir. 2003). “The party offering the expert must prove each of these requirements by a preponderance of the evidence.” *Mahmood v. Narciso*, 529 Fed. Appx. 99, 102 (3d Cir. 2013) (citing *In re TMI Litig.*, 193 F.3d 613, 663 (3d Cir.1999)).

Finally, that courts have traditionally accepted firearms examination testimony sometimes uncritically and without limitation should not be controlling today. The most recent published district court decision in the Third Circuit to address the reliability of firearm toolmark evidence is *United States v. Otero*, 849 F. Supp. 2d 425 (D.N.J. 2012). The Court in *Otero* considered the National Research Council’s 2009 Report’s findings that claims of absolute certainty as to identifications made by practitioners in this area may be overblown and that error rate studies were limited; however, the Court did not yet have the benefit of the subsequent 2016 PCAST Report that found no foundational validity for the discipline.

Courts across the country have reexamined case law in light of the contemporary

understanding of the “science” underlying the method and rejected legal precedent as a basis for admission. *See, e.g., United States v. Adams*, 444 F. Supp. 3d 1248 (D. Or. 2020) (recognizing that the determination that particular shell casings were fired from a particular gun could not be replicated or tested weighed heavily against reliability; that error rates as high as 2.2 percent, allowing a wrongful conviction rate of 1 in 46, were far too high weighed against reliability; that the AFTE theory of identification is not “peer reviewed” in the scientific sense; and that the AFTE theory of identification does not enjoy general acceptance in the broader scientific community); *United States v. Tibbs*, No. 2016-CF1-19431, 2019 WL 4359486, *7–8 (D.C. Super. Ct. Sept. 5, 2019) (discussing “the limited persuasive value of existing case law” as it relates to firearms evidence, given the “perfunctory nature of many of these written decisions”). *See also Murray*, 147 A.3d at 758 (“There is no ‘grandfathering’ provision in Rule 702.”). As the PCAST authors noted: “When new facts falsify old assumptions, courts should not be obliged to defer to past precedent: they should look afresh at the scientific issues.” PCAST Report at 144.

In *Murray*, Judge Easterly further noted that the PCAST Report contains recommendations for trial judges performing their gatekeeper role under Rule 702:

When deciding the admissibility of [forensic] expert testimony, . . . judges should take into account the appropriate scientific criteria for assessing scientific validity including: (i) foundational validity, with respect to the requirement under Rule 702(c) that testimony is the product of reliable principles and methods; and (ii) validity as applied, with respect to [the] requirement under Rule 702(d) that an expert has reliably applied the principles and methods to the facts of the case.

147 A.3d at 759 (Easterly, J., concurring) (quoting PCAST Report at 19). In sum, in “a

case involving scientific evidence—or evidence held out as scientific evidence—*evidentiary reliability* will be based on *scientific validity*.” *Murray*, 147 A.3d at 760 (Easterly, J., concurring) (alteration, citations, and quotations omitted). The three reports discussed in more detail below speak precisely to this issue.

**B. THREE COMMITTEES OF EXPERTS FROM THE
SCIENTIFIC COMMUNITY CONDUCTED A
COMPREHENSIVE REVIEW OF THE EMPIRICAL DATA
AND CONCLUDED THAT FIREARMS IDENTIFICATION IS
NOT SCIENTIFICALLY VALID.**

In the past thirteen years, three reports issued by three separate committees of nationally recognized experts, two from the NRC⁷ and one from the PCAST, have concluded that firearms identification evidence lacks scientific validity. Their conclusions were uniform: fundamental assumptions underlying firearms examination have not been proved; the theory of firearms identification is not a scientific theory; the method is subjective; and there is insufficient empirical evidence establishing validity and estimating reliability of firearms examination. *See PCAST Report* at 104, 111–13; *NRC Forensics Report* at 154; *Ballistics Imaging Report* at 3.

Each committee either included or consulted with independent scientists, statisticians, medical examiners, judges, forensic practitioners, lawyers, and professors with expertise in scientific issues. *See PCAST Report* at v–ix; *NRC Forensics Report* at

⁷ Multiple courts have treated the reports of NRC committees as authoritative works for purposes of evaluating the validity of a scientific field. *See United States v. Porter*, 618 A.2d 629, 642 n.24, 643 n.26 (D.C. 1992). *See also Hayes v. State*, 660 So.2d 257, 264 (Fla. 1995); *Commonwealth v. Gaynor*, 820 N.E.2d 233, 265 (Mass. 2005); *Commonwealth v. Blasioli*, 713 A.2d 1117, 1119 n.3 (Pa. 1998); *State v. Tester*, 968 A.2d 895, 906 (Vt. 2009).

v–xiii; Ballistics Imaging Report at v–vii, xi–xvi, 312–322. Each committee also heard testimony from forensic scientists, reviewed nearly every available journal article and study involving firearms examination, and read every article or study submitted by members of the forensic community. *See* PCAST Report at 2, 155–160; NRC Forensics Report at xx, 2–3; Ballistics Imaging Report at xiii–xvi. *See also* President’s Council of Advisors on Science and Technology, AN ADDENDUM TO THE PCAST REPORT ON FORENSIC SCIENCE IN THE CRIMINAL COURTS 2 (2017) [hereinafter PCAST Addendum; *see* Appendix for website]. The committees were chosen for their special competence and with regard for appropriate balance. *See, e.g.,* NRC Forensics Report at xx, 2; Ballistics Imaging Report at “Notice” on Copywrite Page. Each committee had members with expertise in core sciences—physics, chemistry, biology, materials science, engineering, biostatistics, statistics, and medicine. *See, e.g.,* PCAST Report at v–ix; NRC Forensics Report at xx, 2. With trained scientists on each committee, the committees were uniquely qualified to both define the steps necessary to move a hypothesis or observation into a valid, reliable scientific principle or methodology, and to determine whether each forensic discipline reviewed performed these steps.

Importantly, the committees authoring the reports included scientists versed in scientific methodology, rather than firearms examiners whose financial and professional stake is apparent. These multidisciplinary groups of nationally renowned scientists and professionals are the relevant community to determine whether the requisite research and data demonstrate the scientific validity and reliability of a technique or method discipline. To evaluate the admissibility of firearms identification, this Court must consider the

opinions of the broader scientific community, which are collected and reflected in the three reports discussed below.

1. The 2008 NRC Ballistics Imaging Report

In 2008, a committee of materials scientists, statisticians, and other experts, assembled by the NRC and acting at the request of the Department of Justice, issued a report on firearm pattern-matching analysis. The Department of Justice charged the NRC committee with assessing the feasibility and utility of establishing a “national reference ballistic image database . . . that would house images from firings of all newly manufactured or imported firearms.” Ballistics Imaging Report at 1. But the committee recognized that, in order to answer the questions about the feasibility of such a database, it must first determine “whether firearms-related toolmarks are unique: that is, whether a particular set of toolmarks can be shown to have come from one weapon to the exclusion of all others.” *Id.* at 3. Over the course of its foundational literature review, the committee determined that “[t]he validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.” Ballistics Imaging Report at 3, 81. “A significant amount of research,” the committee concluded, “would be needed to scientifically determine the degree to which firearms related toolmarks are unique or even to quantitatively characterize the probability of uniqueness.” *Id.* at 3.

2. The 2009 NRC Forensics Report

One year after the NRC issued the Ballistics Imaging Report, another committee of experts selected by the NAS issued a scathing report on pattern-matching disciplines, and

the authors did not spare firearms identification from their critique. As recommended by Congress, a committee including “members of the forensic science community, members of the legal community, and a diverse group of scientists” authored the report. NRC Forensics Report at 2. The committee’s findings, including those specific to firearms identification, were informed by its extensive review of scientific literature. The committee also consulted with numerous outside scientists and experts, including—among others—Peter Striupaitis, former president of the Association of Firearm and Toolmark Examiners (AFTE). NRC Forensics Report at xii, 2, 307. Following its review, the committee concluded that insufficient evidence existed demonstrating that firearms examiners can validly and reliably conclude that a piece of ammunition was fired from a particular firearm. *See id.* at 7, 154–55.

The NRC Forensics Report explained that for a forensic discipline to qualify as a generally accepted science, it must meet two basic requirements: (1) its underlying theory and methodology must be tested and validated through repeated, controlled studies measuring error rates and associated confidence intervals; and (2) it must employ specific protocols for the interpretation of the evidence to minimize human error and bias. NRC Forensics Report at 112–124. “[T]he law’s admission of and reliance upon forensic evidence in criminal trials” depends on “the extent to which a forensic science discipline is founded on a reliable scientific methodology, leading to accurate analyses of evidence and proper reports of findings,” and “the extent to which practitioners in those forensic science disciplines that rely on human interpretation adopt procedures and performance standards that guard against bias and error.” *Id.* at 111. The NRC Forensic Report noted

that the AFTE documents, which purportedly provide examiners with the “best guidance available for the field of toolmark identification,” “[do] not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.” NRC Forensics Report at 155. Toolmark and firearms analysis “lack[s] . . . a precisely defined process,” and while AFTE has adopted a so-called “theory of identification,” it “does not provide a specific protocol.” *Id.* The field has no specific, empirical data for an examiner to adhere to when deciding that toolmarks have ““sufficient agreement”” to conclude a common source. Instead, it defines “sufficient agreement” as ““when it exceeds the best agreement demonstrated between tool marks known to have been produced by different tools and is consistent with the agreement demonstrated by tool marks known to have been produced by the same tool.”” *Id.* The meaning of ““sufficient agreement”” does not depend on data, but on the examiner’s “own experience.” *Id.*

For testing and validation, the NRC Forensic Report concluded that “[s]ufficient studies have not been done to understand the reliability and repeatability of the methods,” and no repeated, controlled studies exist from which to devise a “statistical foundation for estimation of error rates.” NRC Forensics Report at 154. The NRC Forensic Report’s authors agreed with those who wrote the Ballistic Imaging Report, concluding that while class characteristics can help narrow the field of possible firearms, “not enough is known about the variability between individual tools and guns” for conclusions of individualization. *Id.* The NRC Forensic Report found that the field of toolmarks and firearms identification had not tested and validated its methodology. In addition, it found

that the field did not employ specific and objective protocols. Finally, the NRC Forensic Report strongly criticized the field's use of an entirely subjective methodology and its failure to require procedures minimizing human error and bias.

3. The 2016 PCAST Report

In 2016, the President's Council of Advisors on Science and Technology issued a report to the President, the PCAST Report. The authors were an advisory group of the nation's leading scientists and engineers and included the President of the Broad Institute (the Broad Institute is a partner of Harvard University and the Massachusetts Institute of Technology), as well as experts in biology, physics, electrical engineering, computer science, and nanotechnology. PCAST Report at vii.

Specifically, in 2015 President Barack Obama requested the PCAST group to determine whether there were "additional steps on the scientific side," in addition to those already taken in response to the "highly critical" 2009 NRC Forensics Report, "to help ensure the validity of forensic evidence used in the Nation's legal system." PCAST Report at x. In contrast to the NRC Forensics Report, which touched on twelve separate disciplines, PCAST examined just six forensic-feature comparison disciplines: firearms analysis; DNA analysis of single-source samples, simple-mixture samples, and complex-mixture samples; bitemark analysis; latent-fingerprint analysis; footwear analysis; and hair analysis. *See id.* at xii. The group's goal was to determine whether those disciplines were scientifically valid and whether they had a methodology that could be reliably applied, which, of course, are the foundational requirements for admissibility of scientific evidence under Rule 702. *See id.* at x.

The group evaluated “a set of over 2,000 papers” and studies from “various sources,” including papers submitted in response to PCAST’s “request for information from the forensic-science stakeholder community.” PCAST Report at 2. It consulted with forensic scientists, including those at the Federal Bureau of Investigation and the National Institute of Standards and Technology. *See id.* Much like in the NRC Forensics Report, PCAST asked whether each forensic discipline meets two key requirements for scientific validity. Specifically, PCAST investigated “foundational validity”—whether the method can, in principle, be reliably applied—and “validity as applied”—whether the method has been reliably applied in practice. PCAST Report at 56.

To be “foundationally valid,” a field must utilize a method that has been subject to “*empirical* testing by multiple groups, under conditions appropriate to its intended use.” PCAST Report at 5. PCAST acknowledged that it is possible for a field to be scientifically valid even if it does not employ an objective method like simple DNA analysis. But the field must still show through studies that the method is “repeatable and reproducible.” *Id.* at 47. A method is “repeatable” if, with a known probability, an examiner can reach the same result when examining samples from the same sources. *Id.* A method is “reproducible” if, with known probability, different examiners can obtain the same result when evaluating the same samples. *Id.*

Put differently, a method is foundationally valid if studies show it has a “reproducible and consistent procedure for: (1) identifying features within evidence samples; (2) comparing the features in two samples; and (3) determining, based on the similarity between the features in two samples, whether the samples should be declared to

be a proposed identification (‘matching rule’).” PCAST Report at 48. The studies must also provide “valid estimates of the method’s accuracy,” demonstrating how often an examiner is likely to draw the wrong conclusions even when applying the method correctly.⁸ *Id.* “Without appropriate estimates of [the method’s] accuracy, an examiner’s statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact.” *Id.* at 6; *see id.* at 48 (“Without an appropriate estimate of its accuracy, a metrological⁹ method is useless—because one has no idea how to interpret its results.”).¹⁰ In other words, in order to be foundationally valid, the field has to “show its work” through black box studies¹¹ documenting that examiners are able to do what they say they can do, and how often.

The second requirement for scientific acceptance, “validity as applied,” requires that the method or technique be “reliably applied in practice.” PCAST Report at 4–5. An

⁸ This is a determination of the method’s accuracy. It is not an estimation of an individual’s accuracy or error rate. For foundational validity, the concern is with estimating the method’s error rate empirically by testing many examiners’ applying the method as specified.

⁹ Metrology is the science of measurement and its application. PCAST Report at 23. All forensic pattern matching disciplines, including firearms analysis, fall under this umbrella.

¹⁰ For objective methods—such as the interpretation of single source DNA evidence—the field can show foundational validity by studying and “measuring the accuracy, reproducibility, and consistency of each of its individual steps” in interpretation. PCAST Report at 5. For subjective feature-comparison methods such as toolmarks or fingerprint analysis, which to date rely only on an examiner’s eyeball comparison of features on known and unknown samples, the method must be evaluated “as if it were a ‘black box’ in the examiner’s head,” with many studies involving numerous examiners “render[ing] decisions about many independent tests” with corresponding error rates determined. *Id.* at 5–6.

¹¹ A black-box study is an empirical study used to estimate a subjective method’s error rate. Black-box studies have significant numbers of examiners analyze samples and render opinions about the origin or similarity of samples. *See* PCAST Report at 48.

examiner must be capable of reliably applying the method, and he or she must have reliably applied the method. To ensure that the examiner is capable of applying the technique, the examiner must conduct proficiency tests that are designed to evaluate how often the examiner reaches the correct answer in circumstances modeling the procedures actually used in case work.¹² *See id.* at 56. To show that the examiner has actually applied the method reliably in each case, the examiner must make available “for scientific review by others” all procedures used, “the results obtained,” and any “laboratory notes” taken. *Id.*

PCAST found that the field of firearms identification failed the test for foundational validity. The PCAST Report, like the NRC Forensics Report and the Ballistics Imaging Report before it, criticized the field of firearms identification’s so-called “theory of identification,” a “theory” based on assumptions rather than scientific data on the frequency of toolmark characteristics or an “empirical demonstration of accuracy.” PCAST Report at 59. A “scientific theory,” PCAST explained, is a “comprehensive explanation of some aspect of nature that is supported *by a vast body of evidence.*” *Id.* at 60 (emphasis added). The “theory” promoted in the toolmark field is not a scientific theory, but rather “a claim that examiners applying a subjective approach can accurately individualize the origin of a toolmark,” a claim without any empirical demonstration of accuracy. *Id.* at 60; *see id.* at 106–112.

¹² Proficiency tests are designed to determine the ability on an individual examiner and not to determine the method’s error rate. Proficiency tests cannot be generalized to establish the method’s error rate; that requires black-box studies.

The PCAST Report stressed that the field must prove scientific validity through rigorous empirical studies, and not through unsupported claims that an examiner can rely on his or her training or experience. “[N]either experience, nor judgment, nor good professional practices . . . can substitute for actual evidence of foundational validity and reliability.” *Id.* at 6. Similarly, “[t]he frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of ‘judgment.’ It is an empirical matter for which only empirical evidence is relevant.” *Id.* Neither can “an expert’s expression of *confidence* based on personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field . . . substitute for error rates estimated from relevant studies.” *Id.* at 55.

Not surprisingly, because the method relies solely on training and experience and not data driven standards, the “method is circular.” PCAST Report at 60. That is, the method “declares that an examiner may state that two toolmarks have a ‘common origin’ when their features are in ‘sufficient agreement,’” but it “defines ‘sufficient agreement’ as occurring when the examiner considers it a ‘practical impossibility’ that the toolmarks have different origins.” *Id.* PCAST raised this concern with the FBI Laboratory, and the lab responded that “[p]ractical impossibility’ is the certitude that exists when there is sufficient agreement in the quality and quantity of individual characteristics. . . .” *Id.* at 60. PCAST rejected this response, explaining that the “practical impossibility” language

did “not resolve the circularity.” *Id.*¹³

The PCAST Report was unequivocal: experience, judgment, and years of use in court cannot establish scientific validity and a degree of reliability. “The *only* way to establish the scientific validity and degree of reliability of a subjective forensic feature comparison method—that is, one involving significant human judgment—is to test it *empirically* by seeing how often examiners actually get the right answer.” PCAST

Addendum at 1. Stated otherwise:

If practitioners of a subjective forensic feature comparison method claim that, through a procedure involving substantial human judgment, they can determine with reasonable accuracy whether a particular type of evidence came from a particular source (*e.g.*, a specific type of pistol or a specific pistol), the claim cannot be considered scientifically valid and reliable until one has tested it by (i) providing an adequate number of examiners with an adequate number of test problems that resemble those found in forensic practice and (ii) determining whether they get the right answer with acceptable frequency for the intended application.

Id. at 1–2. At bottom, scientists must conduct “appropriately designed, black box studies.” PCAST Report at 112.

In contrast to firearms, PCAST deemed latent print examination as a foundationally valid methodology. PCAST Report at 101. PCAST’s criticism of firearms examination therefore cannot be dismissed as setting unreasonably rigorous standards for scientifically sound studies. Given latent print and firearms analyses are both subjective pattern matching methodologies that purport to narrow the pool of potential sources to a single

¹³ See also Peter J. Blau & William A. Tobin, *Hypothesis Testing of the Critical Underlying Premise of Discernible Uniqueness in Firearms-Toolmarks Forensic Practice*, 53 *Jurimetrics J.* 121 (2013) (launching similar critiques).

source, the limitations that have been described for the latent print discipline are informative here. There is no question that the fingerprinting discipline has made greater efforts to establish the reliability and accuracy of its methodology than has the firearms discipline.¹⁴ Still, scientists—including forensic scientists—find that the state of the science is not such that a latent print examiner should be permitted to “identify” a particular person as the source of a latent print.

The American Association for the Advancement of Science (AAAS)¹⁵ weighed in on this issue in the latent print context in 2017, with a report on its multi-year gap analysis of the scientific literature. The report, titled *Forensic Science Assessments: A Quality and Gap Analysis of Latent Fingerprint Analysis* (“AAAS Report”), was authored by a working group comprised of a forensic scientist, a statistician, a biometric engineer, and a psychologist who focuses on cognitive bias issues.¹⁶ In no uncertain terms, this report states that there is “no scientific basis for estimating the number of people who could not be excluded [as the source of a latent print] and, consequently, no

¹⁴ See PCAST Report at 87-88 (“In response to the 2009 NRC report, the latent print analysis field has made progress in recognizing the need to perform empirical studies to assess foundational validity and measure reliability. Much credit goes to the FBI Laboratory, which has led the way in performing both black-box studies, designed to measure reliability, and “white-box studies,” designed to understand the factors that affect examiners’ decisions.”).

¹⁵ AAAS is the world’s largest multidisciplinary scientific society and publisher of the well-respected journal *Science*. “The formation of AAAS in 1848 marked the emergence of a national scientific community in the United States. AAAS was the first permanent organization formed to promote the development of science and engineering at the national level and to represent the interests of all its disciplines.” HISTORY OF AAAS, *available at* <https://www.aaas.org/about/mission-and-history> (last visited Jan. 8, 2021).

¹⁶ *Available at* <https://www.aaas.org/report/latent-fingerprint-examination> (last visited Jan. 8, 2021).

scientific basis for determining when the pool of possible sources is limited to a single person.” AAAS Report at 9.

The AAAS report specifically takes on the language used by the examiner in this case, which parallels SWGFAST language.¹⁷ It finds that “this approach fails to deal forthrightly with the uncertainty that currently exists about the rarity of any given friction ridge impression” given “there is no scientific basis for determining how many people would not be excluded [as the source of a latent print] and no way to determine when the pool of possible sources is limited to a single person.” AAAS Report at 61. In fact, because of this uncertainty, “a latent print examiner has no more basis for concluding that the pool of possible sources is *probably* limited to a single person than for concluding that it is *certainly* limited to a single person.” *Id.* (emphasis in original).

The AAAS Report also specifically addresses an alternate Department of Justice proposal that latent print examiners testify that they “would not expect to see that same arrangement of features repeated in another source,” language that is much more restrained than the language the examiner used in this case. *Id.* As the authors explain, this approach permits examiners “to make an assertion that is deemed scientifically

¹⁷ SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology) is a working group of latent print examiners put together by the FBI. Their definition of “identification,” adopted in 2013, is:

the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source. Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.

AAAS Report at 61, quoting SWGFAST Doc. #10.

unsupportable and improper so long as they hedge by saying that they ‘expect’ that the assertion is true rather than saying outright that it is true.” *Id.* at 62. Moreover, this expectation is informed by “speculation and guesswork, rather than empirical evidence.” *Id.* Because “there is no scientific evidence—none whatsoever—that latent print examiners have the ability to estimate with the required level of precision the frequency of the feature sets observable in latents in the population[. . .] there is no reason to believe that conclusions resting on their ability to make such precise judgments will be reliable or valid.” *Id.* at 63.

“Validation of a method that purports to limit the pool of potential sources to a single firearm would require clearly defined and documented matching criteria, and studies showing that these matching criteria reliably produce an identification to a particular firearm and no others.” *Id.* at 3; *see* K. Kafadar, *Statistical Issues in Assessing Forensic Evidence*, Int’l Stat. Rev. 83(1) (2015) (same). Neither the latent print discipline nor the firearms discipline have conducted these kinds of studies.¹⁸

In contrast to the efforts made in fingerprint comparisons, however, there exists only *one* empirical study attempting to replicate the circumstances of actual case work of firearms examination. PCAST Report at 112. That lone study—the Ames Laboratory Study—hardly establishes that firearms examiners can reliably and repeatedly reach

¹⁸ The OSAC Research Needs document is available online at <https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/osac-research-and-development-needs> (last visited Jan. 8, 2021).

accurate conclusions.¹⁹ In that study, 218 analysts received fifteen separate comparison problems, “each consisting of one questioned sample and three test fires from the same known gun. . . .” *Id.* at 110.²⁰ Those test fires may or may not have been the source of the spent ammunition. “Unbeknownst to the examiners,” each received five same-source and ten different-source comparisons. *Id.* Of the 2,178 different-source comparisons, analysts found 1,421 eliminations, 735 inconclusives, and 22 false positives. *Id.*

These results translate to a false positive rate of 1.5%, with a 95% confidence bound placing the error rate as high as 2.2%.²¹ PCAST Report at 110. That false positive rate correlates to an estimated error rate between one in forty-six cases and one in sixty-six cases—a far cry from the “near-perfect accuracy” often claimed by examiners. *Id.* at 105, 110. Equally problematic, the examiners had a 33.7% inconclusive rate. *Id.* As such,

¹⁹ This is the only study performed by researchers not associated with a crime lab. Rather, Ames Laboratory is a Department of Energy national laboratory affiliated with Iowa State University. PCAST Report 109–110.

²⁰ Forensic practitioners themselves raise the criticism that this study did not involve consecutively manufactured firearms. PCAST Report at 331. Given that it is generally considered more difficult to distinguish consecutively manufactured firearms (*e.g.*, because they are more likely to possess subclass marks that can be mistaken for “individual” marks), this suggests that the true error rate could be significantly higher than this single estimate.

²¹ These error rates are appropriately based on conclusive examinations (*i.e.*, inconclusive results are not part of the total). “[F]rom a scientific standpoint, [false positive error rate based on conclusive examinations] should be used for reporting FPR to a jury. This is appropriate because evidence used against a defendant will typically be based on *conclusive*, rather than inconclusive, examinations.” PCAST Report at 153. For example, for a method that produces 990 inconclusive results and ten false positives in 1,000 examinations, it would be misleading to report that false positive rate as 1% (10/1000 examinations). “Rather, one should report that 100% of the conclusive results were false positives.” *Id.* Further, as false positive rates can only be estimated, “it is necessary and appropriate to quote confidence bounds within which [the false positive rate is] highly likely to lie. Because one should be primarily concerned about . . . underestimating FPR, it is appropriate to use a *one-sided* confidence bound. Upper 95 percent one-sided confidence bounds should thus be used for assessing error rates [and] [t]he use of lower values may rightly be viewed with suspicion as an attempt at obfuscation.” PCAST Report at 152–53.

the “science” is not developed to a point that would allow examiners to reliably inculcate or exculpate.

Unsurprisingly, PCAST concluded that firearms identification did not meet the criteria for foundational validity based on this sole study. “[T]here is only a single study that was appropriately designed to measure validity and estimate reliability.” *Id.* at 111. In addition, the Ames Study, although available on the web, had “not yet been subjected to peer review and publication.” *Id.* “The scientific criteria for foundational validity,” PCAST concluded, require “appropriately designed studies by *more than one group* to ensure reproducibility.” *Id.* “Because there has been only a single appropriately designed study, the current evidence falls short of the scientific criteria for foundational validity.” *Id.*

The core agreement across these three independent scientific reports—the 2008 NRC Ballistics Imaging Report, the 2009 NRC Forensics Report, and the 2016 PCAST Report—is that the firearms toolmarks comparison discipline lacks foundation validity.

**C. THIS COURT SHOULD EXCLUDE FIREARMS
IDENTIFICATION EVIDENCE AS SCIENTIFICALLY
INVALID.**

The conclusions of the three reports provide powerful evidence that such evidence is properly excluded under Rule 702: the field of firearms identification lacks the scientific validity that is the predicate for evidentiary admissibility. Indeed, the conclusions of these reports are uniform: There is no scientific theory or technique that is generally accepted by the relevant scientific community, and there is no evidence that examiners regularly reach accurate conclusions. *See generally Daubert*, 509 U.S. at 593–

95. This Court should take note of these consistent findings and exclude Analyst Mills' proffered testimony.

Each of the three reports discussed above found no empirical research proving the “theory of identification.” Instead, this so-called “theory” is nothing more than a series of unproven assumptions embraced by firearms examiners. First, the field of toolmarks and firearms examination assumes that every firearm has unique characteristics resulting from both the manufacturing process and wear and tear, which consistently produce unique marks on bullets and shell casings. Second, the field assumes that examiners, based on experience and judgment, can identify and distinguish these “unique” accidental marks from subclass marks, which are also accidentally left during the manufacturing process but are not unique. Third and finally, the field assumes that examiners can accurately determine when two sets of “unique” marks are in “sufficient agreement” to declare a “match.” There is no standard for what constitutes “sufficient agreement,” and there is no empirical data supporting any particular threshold number of marks that must align. Rather, each examiner refers to his or her personal recollection of the amount of agreement he or she has observed in close non-matches encountered throughout her career; in other words, “sufficient agreement” is wholly subjective.

Assumptions and anecdotal observations do not constitute a valid scientific theory tested through independent empirical research. To the contrary, a valid “scientific theory” is a “comprehensive explanation of some aspect of nature that is supported by a vast body of evidence.” PCAST Report at 60; *see* NRC Forensics Report at 154–55 (reaching same conclusion); Ballistics Imaging Report at 1, 3 (same).

These reports also conclude that the field of firearms identification has not proved it has a reliable or valid scientific technique for reaching conclusions about “individualization.” As all three reports recognized, to be an actual scientific method, there must be empirical testing to prove that examiners can reliably and repeatedly reach accurate conclusions. The field of firearms identification has not engaged in such testing, and firearms analysis currently “falls short of the scientific criteria for foundational validity.” PCAST Report at 111.

The field has insufficient empirical data showing that examiners accurately do what Analyst Mills’ apparently claims to have done in this case—reliably match pieces of ballistic evidence as having been fired from a specific firearm. Nor does the field have any repeatable and consistent procedures for examiners to follow when identifying features within evidence samples, comparing those samples, and then measuring the significances of similarities and differences. Rather, the field allows each examiner to draw his or her own subjective conclusions about what qualifies as “sufficient agreement” between test-fired samples and found ammunition, or between sets of found ammunition. *Id.* at 111–12; *see* NRC Forensics Report at 155 (concluding that field of toolmarks relies on the “subjective findings of examiners rather than on the rigorous quantification and analysis of the sources of variability.”); *id.* at 153–54 (noting that “the decision of the toolmark examiner [to declare a match] remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates”).

Equally problematic, the field lacks the rigorous proficiency testing and controls

for minimizing human error and bias that ensure scientific validity as applied. Whereas “[s]cience takes great pains to avoid biases by using strict protocols to minimize their efforts,” NRC Forensics Report at 122, the field of firearms identification has not taken any steps in that regard. *Id.* at 8 n.8 (“Unfortunately, . . . there is no good evidence to indicate that the forensic science community has made sufficient effort to address the bias issue; thus, it is impossible for the committee to fully assess the magnitude of the problem.”). To the contrary, firearms examiners are often given contextual information about the case and evidence before performing examinations. *See id.* at 123. And the field lacks the rigorous proficiency testing showing that specific examiners are capable of conducting accurate evaluations in each case. *See* PCAST Report at 111–12.

As these reports also make clear, the field lacks sufficient peer review. *See generally Daubert*, 509 U.S. at 593–95 (discussing importance of presence (or absence) of peer-reviewed studies demonstrating field’s validity). Only one study, the Ames Study, has replicated case-work situations to test the reliability of case work. This does not establish scientific validity. *See* PCAST Report at 111.

And finally, the field lacks a known error rate, as only one appropriately designed study—the Ames Study—has attempted to measure it. PCAST Report at 108–110; *see generally Daubert*, 509 U.S. at 593–95 (emphasizing importance of known error rate in assessing field’s reliability). The Ames Study suggests that the field has grossly underestimated its rate of false-positives. *See* PCAST Report at 11, 111. Without a known error rate, “an examiner’s statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and

considerable potential for prejudicial impact.” *Id.* at 6.

Therefore, even if the government were to respond with a common assurance that it will not elicit identifications of a cartridge case or a bullet to a firearm to the exclusion of all other firearms, this provides nothing more than a false sense of security. As Judge Easterly points out in her concurrence in *Williams v. United States*, 130 A.3d 343 (D.C. 2016), an identification implicitly communicates the same (mis)information to the jury.²²

The government cannot show that the field of firearms examination has established—through empirical research, rather than unsupported assertions—that its underlying theory is true, that toolmarks are unique, that an examiner can follow a proven methodology to declare a match, and that its examiners produce accurate results when applying that methodology. In light of this, the Court should exclude Analyst Mills’ proposed testimony.

**D. THE COURT SHOULD EXCLUDE ANALYST MILLS’
TESTIMONY BECAUSE THERE IS NO WAY TO DISCERN
WHETHER HIS RESULTS ARE RELIABLE.**

Even if the general methodology of firearm toolmark comparison survives the critiques leveled at it by three separate groups of independent scientists, this Court must query “whether those principles and methods have been properly applied to the facts of the case.” *See* Fed. R. Evid. 702 advisory committee note; Fed. R. Evid. 702(d). The

²² “If purportedly unique patterns on bullets are declared a match, that declaration likewise negates the possibility that more than one gun could have fired the bullets—it is a statement of unqualified certainty that the bullets were fired from a specific gun to the exclusion of all others.” *Williams*, 130 A.3d at 351 (Easterly, J., concurring).

AFTE theory of identification supplies the barest conceptual framework for firearm toolmark comparison, and it is essential that firearms examiners and the laboratories at which they practice flesh out this “theory” to produce reliable results in practice. This is particularly true where, as here, an examiner concludes that a particular firearm is the source of marks on the recovered ammunition, and in so doing determines that “agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.” AFTE THEORY OF IDENTIFICATION, available at <https://afte.org/about-us/what-is-afte/afte-theory-of-identification> (last visited Jan. 8, 2021).

This is an extremely heavy load to bear, and pursuant to Rule 702(d), this Court should expect to see thorough documentation supporting and bearing on the reliability of each analytical step required to reach such a conclusion. *Cf. United States v. Davis*, No. 4:18-cr-00011, 2019 WL 4306971, *8 (W.D. Va. 2019) (ordering the government to produce pursuant to Fed. R. Evid. 702 and Fed. R. Crim. P. 16 (a)(1)(G) a “sufficient narrative to provide the bases and reasons for the opinions reached by them as to the presence of similar or consistent toolmarks”). “[A]ny step that renders the analysis unreliable . . . renders the expert testimony inadmissible. *This is true whether the step completely changes a reliable methodology or merely misapplies that methodology.*” Fed. R. Evid. 702 advisory committee note (quoting *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 745 (3d Cir. 1994)). As-applied validity in the firearms context requires evidence that (1) the examiner, an inextricable part of any subjective pattern matching discipline, is capable of reliable analysis on sufficiently challenging samples to ensure proficiency; (2)

the protocols that the examiner applies and the documentation produced are sufficient to ensure reliable and repeatable analysis;²³ and (3) those protocols were in fact followed. The government is unable to make such a showing in this case.

1. Proficiency Testing is Insufficiently Rigorous to Ensure Reliability.

Because firearms examination “depends on subjective judgment, it is scientifically unjustified to conclude that a particular examiner is capable of reliably applying the method unless the examiner has undergone regular and rigorous proficiency testing.” PCAST Report at 101 (referring to latent print analysis); *see id.* at 113 (“validity as applied [in firearms analysis] would . . . require that the expert has undergone rigorous proficiency testing on a large number of test problems to evaluate his or her capability and performance”). However, a representative of Collaborative Testing Services (CTS), a proficiency test provider serving the forensic firearms community, has publicly disclosed that a driving force in its test design is its customers’ preference for easy tests.²⁴ Firearms examiners themselves concede that CTS proficiency tests are “insufficiently challenging to be of use in demonstrating competence in comparison microscopy skills.” P. Pauw-

²³ “If a laboratory consistently fails to use certain quality controls so that its results are rendered unreliable, attempting to ascertain whether the lack of quality controls constitutes a failure of methodology or a failure of application of methodology may be an exercise in metaphysics.” *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d at 745.

²⁴ C. Czyryca, A PROFICIENCY TEST PROVIDER’S PERSPECTIVE, COLLABORATIVE TESTING SERVICES, INC., *presented at* August 2015 meeting of the National Commission on Forensic Science, at slide 12, available at <https://www.justice.gov/archives/ncfs/page/file/761061/download> (last visited Jan. 8, 2021) (discussing challenges to proficiency test providers, including “samples without a consensus [*i.e.*, where everyone taking the test does not get the same result] are viewed as poor quality, even if reproduction is true to original” and noting that “[e]asy tests are favored by the community”).

Vugts, et al, *FAID2009: Proficiency Test and Workshop*, AFTE J. 45(2): 115–127, 117 (2013). Upon information and belief, virtually every proficiency test taken by the examiner involved in this case over the course of his career was issued by CTS. Thus, there is no way to assess his reliability with samples representative of casework.

The chasm between these easy proficiency tests and proficiency tests that include “realistic and challenging comparisons” is illustrated by the findings of a large-scale proficiency test (“FAID2009”) issued in response to the recognized need for more rigorous testing.²⁵ Pauw-Vugts, et al, *supra*, at 117–18. FAID2009 involved ten test sets of fired ammunition produced using firearms that were recovered during casework in the United States, Norway, and the Netherlands. *Id.* at 118. The results from this proficiency test “show a disturbingly high amount of false identifications and exclusions.” *Id.* at 124. The high error rate with rigorous proficiency testing underscores the negligible value of much easier proficiency tests like those provided by CTS in establishing competency under casework-type conditions.²⁶

²⁵ FAID2009 is the second in a series of collaborative proficiency tests; the first test is known as FAID2005. Pauw-Vugts et al., *supra*, at 117. In FAID2005, as with FAID2009, “[t]he biggest cause of errors . . . was found to be the failure to recognize sub-class breechface marks.” *Id.*

²⁶ The results of the FAID2009 test also support PCAST’s finding that the consecutively manufactured firearm studies, in which the firearms community puts so much stock in establishing the validity of their discipline, suffer from faulty design that makes them unrealistically easy and their error rates irrelevant to caseworking scenarios. PCAST Report at 107–08. Among other things, participants in those studies were aware that they were a part of a consecutive firearm study, and thus were primed to be on the lookout for subclass marks that might otherwise be mistaken for individual marks. With the FAID2009 test, participants were not aware that such close non-matches were part of the test set (neither would an examiner have such knowledge in a casework scenario), and a quarter of them overvalued similarities and dismissed differences observed between casings fired from different guns. Pauw-Vugts et al, *supra*, at 126

2. **Without further discovery of the litigation packets, it is impossible to determine (i) whether any protocols that the examiner applied and the documentation produced are sufficient to ensure reliable and repeatable analyses and (ii) whether those protocols were in fact followed.**

Where, as here, a case involves technical and scientific evidence, detailed expert disclosure of litigation packets is essential to a defendant's right to bring a *Daubert* challenge and to the right to cross-examine a government toolmark expert at any *Daubert* hearing or trial. *See United States v. Jackson*, 51 F.3d 646, 651 (7th Cir. 1995) (“[C]ases involving technical or scientific evidence [] may require greater disclosure, including written and oral reports, tests, investigations, and any other information that may be recognized as a legitimate basis for an opinion under Fed. R. Evid. 703.”). Counsel cannot adequately cross-examine the prosecution's firearm expert or prepare a complete *Daubert* challenge without understanding the scientific bases, if any, underlying the conclusory firearm examination report provided in this case. *See, e.g., United States v. Caputo*, 382 F. Supp. 2d 1045 (N.D. Ill. 2005) (“It is exceedingly difficult to cross-examine a scientific expert witness about the results of a scientific test without an opportunity to first review the test giving rise to the results.”); *United States v. Robinson*, 44 F. Supp. 2d 1345, 1346 (N.D. Ga. 1997) (in a fingerprint case, Rule 16(a)(1)(G) required disclosure of all the points of identification on which the government's expert would rely as a basis for her opinion that the defendant's prints appeared on evidence; “[i]f a defendant does not have the bases for the government's opinion, there is no way the defendant's counsel can effectively cross-examine the expert. It is this issue, which

goes to the fairness of the trial, that the court must always keep in mind in dealing with discovery issues in a criminal case.”).

Thus, as *Robinson* holds, merely providing a conclusory report, or even providing photographic comparison data that are merely side-by-side comparisons, does not suffice; the specific points of identification being relied upon to make an identification must be documented and disclosed. As *Robinson* also holds, the government’s failure to comply with this obligation is alone sufficient to exclude the testimony of an expert under Rule 16. *Accord United States v. Saunders*, 826 F.3d 363, 369 (7th Cir. 2016) (citing *Robinson*) (“We agree with the defendants that the government’s Rule 16 disclosure was insufficient because it failed to provide the number of points of identification that were the basis for Ambrozich’s opinion that the fingerprints were a match.”). As the Seventh Circuit underscored in *Saunders*, “[w]ithout knowing the . . . points used by the government’s expert in advance, a defendant’s ability to prepare an attack on the validity of the identification may be hindered.” *Id.* at 370.

The same logic holds true in firearm examination cases. *See United States v. Willock*, 696 F. Supp. 2d 536, 570 (D. Md. 2010) (“To ensure that defense counsel can make any challenges to the admissibility of toolmark identification evidence and that courts may conduct hearings to resolve these challenges based on sufficient record, the Government should be required to strictly and timely comply with its Fed.R.Crim.P. 16 obligations regarding the opinions to be offered by firearms examiners in sufficient detail and sufficiently far in advance of motions deadlines or trials as to enable defense counsel to evaluate the conclusions and bases, determine whether to engage experts to test them,

and if appropriate, challenge them.”); *United States v. Monteiro*, 407 F. Supp.2d 351 (D. Mass. 2006) (ballistics expert testimony was inadmissible because examiner failed to document the basis for his notation “positive ID” with any photograph or other documentation and thus failed to comport with the field’s own standards for documentation.); *United States v. Green*, 405 F. Supp. 2d 104, 108 (D. Mass. 2005) (firearm examiner’s ultimate opinion excluded, in part because he “took no notes, recorded no measurements, made no photographs, and drew no diagrams”); *Commonwealth v. Pytou Heang*, 458 Mass. 827, 847 (Mass. 2011) (“[B]efore trial, the examiner must adequately document the findings or observations that support the examiner’s ultimate opinion, and this documentary evidence, whether in the form of measurements, notes, sketches, or photographs, shall be provided in discovery, so that defense counsel will have an adequate and informed basis to cross-examine the forensic ballistics expert at trial.”).

In this case, the conclusory report provided by the government does not reveal what methodology the examiner used in making his comparisons, and it does not document what points of comparison the expert is relying upon to make his identifications. This information would typically be revealed in the litigation packets that the defense has previously requested. In order to ensure that Mr. Bowers has a full and fair opportunity to raise all issues in support of the present *Daubert* motion, this Court should order the government to produce the litigation packets forthwith. As Judge Gertner concluded in *Green*:

While I recognize that the *Daubert-Kumho* standard does not require the illusory perfection of a television show (CSI, this wasn't), when liberty hangs in the balance—and, in the case of the defendants facing the death penalty, life itself—the standards should be higher than were met in this case, and than have been imposed across the country. The more courts admit this type of toolmark evidence without requiring documentation, proficiency testing, or evidence of reliability, the more sloppy practices will endure; we should require more.

Green, 405 F. Supp. 2d at 109.

**E. THE PREJUDICIAL IMPACT OF FIREARMS ANALYSIS
EVIDENCE SUBSTANTIALLY OUTWEIGHS ANY
PROBATIVE VALUE.**

As discussed above, the probative value of an “identification,” “match,” or “association” between a spent cartridge or shell casing and a gun is entirely speculative. But even if this Court were to determine that there might be some value in toolmark analysis, it is plain that the prejudicial impact of such evidence substantially outweighs any potential probative value and should be excluded on that ground. *See* Fed. R. Evid. 403 (“Rule 403”). Even assuming it is possible to include a particular weapon as a potential source of a projectile, how many other weapons might also be “included” is unknown. Thus, the jury will have no way of knowing whether one in ten guns could have produced a particular stria on the bullet, or one in ten thousand, or one in ten million, or one in ten quadrillion guns. The potential that they will conclude, without any support, that it is the first (one in ten) is simply too dangerous to permit. *See United States v. St. Gerard*, APO AE 09107, at 4 (U.S. Army Trial Judiciary, 5th Jud. Cir., June 7, 2010) (“[T]he probative value of [the expert’s] proffered testimony that it would be practically impossible for a tool other than the seized AK-47 to have made the marks on

the cartridge case would be substantially outweighed by the unfair prejudice associated with its unreliability.”), http://clpex.com/swgfast/Resources/101126_US-v-Gerard.pdf (last visited Jan. 8, 2021).

This danger is especially heightened where, as here, the testimony to be offered is alleged to be “scientific.” Courts have long recognized that such “scientific” or “expert” evidence can carry great but unwarranted weight with jurors. *See Daubert*, 509 U.S. at 595 (“Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it.”); *United States v. Frazier*, 387 F.3d 1244, 1263 (11th Cir. 2004) (“Simply put, expert testimony may be assigned talismanic significance in the eyes of lay jurors, and, therefore, the . . . courts must take care to weigh the value of such evidence against its potential to mislead or confuse.”). Scholars, too, have recognized that jurors are, even under the best of circumstances, likely to overvalue forensic evidence. *See, e.g.*, Keith A. Findley, *Innocents at Risk: Adversary Imbalance, Forensic Science, and the Search for Truth*, 38 Seton Hall L. Rev. 893, 948 (2008) (“[R]esearch indicates that jurors often do not understand the fundamentals of scientific evidence, and lack the ability to reason about statistical, probabilistic, and methodological issues effectively.”); Mark A. Godsey & Marie Alao, *She Blinded Me with Science: Wrongful Convictions and the “Reverse CSI Effect,”* 17 Tex. Wesleyan L. Rev. 481, 495 (2011) (noting that “jurors in this country often accept state forensic testimony as if each prosecution expert witness is the NASA scientist who first put man on the moon”); Tom R. Tyler, *Viewing CSI and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction*, 115 Yale L.J. 1050, 1068 (2006) (“There is widespread evidence indicating that people already

overestimate the probative value of scientific evidence.”); Richard H. Underwood, *Evaluating Scientific and Forensic Evidence*, 24 Am. J. Trial Advoc. 149, 166 (2000) (“Given their lack of scientific sophistication and innumeracy, jurors are likely to overestimate the significance of [expert testimony].”).

The value of firearms evidence is, at best, negligible. Given that, and given the serious risk that its introduction would mislead the jury, it must be excluded under Rule 403.

**F. INTRODUCTION OF INDIVIDUALIZATION TESTIMONY
WOULD VIOLATE MR. BOWERS’ DUE PROCESS RIGHTS.**

The introduction of Analyst Mills’ opinion that [REDACTED]

[REDACTED] would not only flout *Daubert* and Rules 702 and 403, but also violate Mr. Bowers’ due process rights under the Constitution to a trial free of false and misleading evidence. Expert testimony purporting to [REDACTED]

[REDACTED] is subjective speculation masquerading as “scientific” evidence, which is, at best, misleading, and false as a matter of science. “[I]t is established that a conviction obtained through use of false evidence, known to be such by representatives of the State, must fall under the Fourteenth Amendment.” *Napue v. Illinois*, 360 U.S. 264, 269 (1959); U.S. CONST. amend. XIV; *see Ege v. Yukins*, 485 F.3d 364 (6th Cir. 2007) (admission of foundation-free expert testimony violated due process); *United States v. Ausby*, 916 F.3d 1089 (D.C. Cir. 2019) (*Napue* violation for introduction

of hair comparison evidence); *State v. Bridges*, No. 90 CRS 23102-04, 2015 WL 12670468, at *2 (N.C. Super. Oct. 1, 2015) (introduction of hair comparison evidence was a due process violation because it “exceeded the limits of the science and overstated the significance of the hair analysis to the jury”).

G. IN THE ALTERNATIVE, THIS COURT SHOULD LIMIT THE SCOPE OF THE EXPERT’S TESTIMONY, REQUIRE ERROR RATE TESTIMONY, AND EXCLUDE ANY CONCLUSIONS UNGROUNDED IN A RELIABLE SCIENTIFIC FOUNDATION.

Should this Court refuse to exclude Agent Mills’ testimony on firearm identification, it should nonetheless prevent him from testifying that [REDACTED]

Instead, the examiner should be limited to testifying that [REDACTED].
[REDACTED].²⁷ Failure to exclude is more in line with the scope of the analysis conducted (a single side-by-side comparison with limited documentation) than identification of a particular firearm. Courts must limit the examiner’s testimony if there is a gap between the conclusions supported by existing data and the examiner’s proposed testimony. *See General Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997); *Sargon Enter. Inc.*

²⁷ The defense does not concede that limiting Agent Mills’ proffered testimony in this way will prevent the expert, and therefore the government, from misleading the jury. Studies suggest that jurors still believe that there is an unqualified match between two sources, even when experts provide limiting language. *See* Dawn McQuiston-Surrett & Michael Saks, *Communicating Opinion Evidence in the Forensic Identification Sciences: Accuracy and Impact*, 59 *Hastings L.J.* 1159, 1188–89 (2008); *see* Sarah Lucy Cooper, *The Collision of Law and Science: American Court Responses to Developments in Forensic Science*, 33 *Pace L. Rev.* 234 (2013). In other words, the suggestion here is an alternative position and in no way should be construed as a forfeiture or waiver of the defense’s primary position that this Court should exclude Analyst Mills’ testimony in its entirety.

v. Univ. of S. Cal., 288 P.3d 1237, 1258 (Cal. 2012). “The frequency with which a particular pattern or set of features will be observed in different samples . . . is not a matter of ‘judgment.’ It is an empirical matter for which only empirical evidence is relevant.” PCAST Report at 6. As explained in the preceding sections, researchers have not yet produced reliable empirical evidence of the frequency of toolmark characteristics.

The NRC 2008 Ballistics Imaging Report was not only critical of the practice of toolmark examinations, but also disapproved of the habit of examiners testifying to their opinions in terms of scientifically unsupportable certainty. In the subsequent 2009 Forensic Sciences Report, the National Academy of Sciences stated that AFTE “does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.” 2009 Forensic Sciences Report at 155. Thus, as was done in *United States v. Willock*, 696 F. Supp. 2d 536, 569–70 (D. Md. 2010), toolmark evidence should only, if ever, be admitted when presented by: (1) a qualified examiner, (a) who followed the AFTE theory, and (b) who documented with notes, photographs, or sketches the conclusions reached in sufficient detail to permit; and (2) confirmation by a second qualified examiner of how an identification was reached. In addition, the examiner must be prevented from making outlandish and unsupported pronouncements about the degree of certainty of his identification. *Id.*

Courts across the country have similarly and increasingly limited expert testimony. *See United States v. Adams*, 444 F. Supp. 3d 1248 (D. Or. 2020) (limiting expert testimony to general class characteristics observed and precluding any evidence of

the examiner's methodology or conclusions relating to whatever shell casings matched a particular firearm); *Davis*, 2019 WL 4306971, at *3 (precluding claims of individualization and any expression of confidence and limiting expert's opinion "to whether toolmarks on certain cartridge cases bear marks consistent with each other"); *Tibbs*, 2019 WL 4359486, at *23 (limiting expert testimony to opinion that the "recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting"); *United States v. Shipp*, 422 F. Supp. 3d 762 (E.D.N.Y. 2019) (same); *United States v. Medley*, 312 F. Supp. 3d 493 (D. Md. 2018) (same); *Green*, 405 F. Supp. 2d 104, 124 (D. Mass. 2005) (citing *United States v. Hines*, 55 F. Supp. 2d 62 (D. Mass. 1999)) (permitting testimony only regarding an examiner's observations *without any* accompanying conclusions about the source of a projectile); *United States v. Love*, No. 2:09-cr-20317-JPM (W.D. Tenn. Feb. 8, 2011) (excluding testimony regarding absolute or practical certainty); *United States v. Alls*, No. 2:08-cr-00223-ALM (S.D. Ohio Dec. 7, 2009) (forbidding any claim of a match to one firearm to the exclusion of all other guns and limiting examiner to descriptions of her methodology and observations of casings); *United States v. Glynn*, 578 F. Supp. 2d 567, 571, 575 (S.D.N.Y. 2008) (noting that, given the lack of data supporting the discipline, "ballistics lacked the rigor of science;" limiting testimony regarding a match to a conclusion of "more likely than not" instead of even "reasonable ballistics certainty" to ensure that "a conviction in a criminal case [] not rest *exclusively* on ballistics testimony"); *Missouri v. Goodwin-Bey*, No. 1531-CR00555-01 (Cir. Ct. Green County, Mo., Dec. 16, 2016)

(limiting testimony “to the point this gun could not be eliminated as the source of the bullet”).

At most, this Court should allow the examiner to testify to the general class characteristics observed. No evidence relating to methodology or conclusions relating to whether shell casings match a particular firearm should be admitted at trial. *See United States v. Adams*, 444 F.Supp.3d 1248, 1267 (D. Or. 2020).

If the examiner is permitted to go beyond general class characteristics, this Court must also require the examiner to accurately report an error rate for firearms comparisons. To date, the only available error rate comes from the Ames Study. Analyst Mills should therefore be required to clarify when reporting his results that the field has an error rate between one in forty-six and one in sixty-six. *See PCAST Report* at 12 (“If firearms analysis is allowed in court, the scientific criteria for validity as applied should be understood to require clearly reporting the error rates seen in appropriately designed black-box studies [the Ames Report]. Claims of higher accuracy are not justified at present.”) Further, because the examiner in this case has not been subjected to rigorous proficiency testing, he should disclose that there is no data upon which to assess his reliability. Without this qualifying information, there is a grave risk that the jury will overvalue the expert’s testimony.

IV. CONCLUSION

WHEREFORE, for the reasons set forth above, and any additional reasons that may appear to this Court or arise pursuant to a hearing on this Motion, Mr. Bowers

respectfully moves this Court to exclude certain testimony, or in the alternative limit the testimony, of Agent Mills regarding firearm toolmark “identification” claims.

Respectfully submitted,

/s/ Judy Clarke

Judy Clarke

Clarke Johnston Thorp & Rice, PC

/s/ Michael J. Novara

Michael J. Novara

First Assistant Federal Public Defender

/s/ Elisa A. Long

Elisa A. Long

Supervisory Assistant Federal Public Defender

Exhibit 1

(Filed Under Seal)